AD-249 662

TECHNICAL PAPER
ORO-TP-15

PRO CON

ORO

OPERATIONS RESEARCH OFFICE

The Johns Hopkins University

A Procedure for Determining a Family of Minimum-Cost Network Flow Patterns

by

Robert G. Busacker
Paul J. Gowen

November 1960

Operating Under Contract with the

DEPARTMENT OF THE ARMY

This paper contains opinions of the author(s) only.

OPERATIONS DIVISION
Technical Paper ORO-TP-15
Published November 1960

A Procedure for Determining a Family of Minimum-Cost Network Flow Patterns

by

Robert G. Busacker
Paul J. Gowen

OPERATIONS RESEARCH OFFICE
The Johns Hopkins University Bethesda, Maryland

Received for Publication
21 October 1960
Published
November 1960
by
The Johns Hopkins University
OPERATIONS RESEARCH OFFICE
6935 Arlington Road
Bethesda, Md.

FOREWORD

The ORO research group to which the authors of this paper are assigned just completed the transportation phase of a study of computer-assisted strategic planning for the Army.* A central difficulty in such planning is analysis of major transportation networks, so that one objective of the research was programing a digital computer to carry out network analysis in a manner conforming with Army requirements. This in turn called for the adaptation or creation of a suitable algorithm.

The solution obtained does the following:

(a) For any existing rail or highway net, determines the maximum tonnage that can be moved from any number of specified "sources" (e.g., ports and beaches) to any number of specified "destinations" (e.g., depots in the Communications Zone of a military theater of operations, or Army supply points).

(b) For any tonnage requirements at the destinations less than or equal to the maximum, determines the routings necessary to deliver such tonnage.

(c) Subject to (a) and/or (b) above, determines the minimum-distance routings to accomplish the required deliveries.

(d) Determines, for each routing in the solution, the tonnage flow over each individual link.

*Reported in ORO-T-393, "Computer-Assisted Strategic Logistic Planning: Transportation Phase", now in publication.

The computer program thus obtainable prints routes, with link and junction point designations, in the form required as input to the subsequent steps of the transportation planning process.

A full discussion of the application and some discussion of the algorithm is being made available in ORO-T-393, to which readers interested in these aspects are referred. It was decided to restrict the present paper entirely to the formal statement of the mathematical problem underlying the algorithm, with its solution and the required proofs.

The aim of the work was to develop an algorithm tailor-made to the Army's requirements. The most directly applicable ideas in the literature were studied and adapted. Thus the results in this paper are not entirely original but constitute in part an adaptation of known methods to a specific new application. However, the <u>ideal solution</u>* yielding a minimum "cost" or "distance" flow pattern for every quantity of flow up to the maximum, with the associated <u>cost profile</u>,* are considered to be novel.

Strother H. Walker, Chairman
Logistics Gaming Group

4 November 1960

*See definitions below, p 22 and p 23 respectively.

ABSTRACT

A procedure is presented for solving the following problem:

> Given a finite network with a capacity and a length (or cost factor) associated with each orientation of every link, find a family of minimum-cost flow patterns between two specified nodes, one pattern for every integral amount of flow up to the maximum consistent with link capacities.

The procedure is an iterative process that adds a succession of appropriate elementary chain flows in such a way that each new pattern minimizes cost for a greater amount of flow. The presentation is intended to be self-contained and includes a proof of the validity of the procedure.

CONTENTS

A PROCEDURE FOR DETERMINING A FAMILY OF MINIMUM-COST NETWORK FLOW PATTERNS

INTRODUCTION

This paper presents a general procedure for solving the following abstract network flow problem: Given a finite network, whose links have assigned capacities and lengths (or cost factors), find a family of minimum cost flow patterns between two specified nodes, one pattern for every integral amount of flow up to the maximum consistent with link capacities.

The first section defines the basic network terms and symbols employed throughout the paper. A number of lemmas that are consequences of these definitions are also established here. The second section presents an iterative procedure that is shown to solve the stated network problem, under the assumption that an algorithm with certain characteristics can be devised. The final section of the paper produces this algorithm and proves that it has the required characteristics. A glossary of the principal symbols appearing in the development is included at the end.

Some of the basic ideas introduced in the paper are illustrated geometrically. However, the mathematical development does not rest on geometrical or topological considerations. It should also be noted that network theory, including the specific topic of this paper, is applicable to problems in which the relevant network is not a physical structure connecting points in space. Whereas network links may be interpreted as transport routes, they may also be interpreted as the set of permissible information channels in an organization or as the set of possible transitions between the various states of a system.

PRELIMINARY DEFINITIONS AND LEMMAS

BASIC NETWORK TERMS

A "network" consists of a finite set of distinct elements $N_1, N_2, \ldots, N_k$ called "nodes", together with a subset of all unordered pairs (N_i, N_j) that can be formed from distinct nodes N_i and N_j. The elements of this subset are termed the "links" of the network. These are necessarily finite in number, an upper limit on the number of links being $[k(k-1)]/2$, where k is the number of nodes. (This is the number of links if every pair of distinct nodes determines a link.)

If an ordering is assigned to the nodes of a link (N_i, N_j) designating N_j as the first or "initial node" and N_i as the second or "terminal node", . . . the resulting object is termed a "directed link" from N_j to N_i and is denoted by $\overline{N_j, N_i}$. Thus two directed links, $\overline{N_i, N_j}$ and $\overline{N_j, N_i}$, can be associated with each link.

It is convenient to introduce a representative set of directed links called an "enumerating set." If a network has n undirected links, an enumerating set for this network is any set of n directed links such that each undirected link is represented by precisely one of the two directed links associated with it. It is evident that there are 2^n distinct enumerating sets that can be associated with a network. The function of such a set is to establish for each link a convention as to which direction will be considered positive in describing flows through the network.

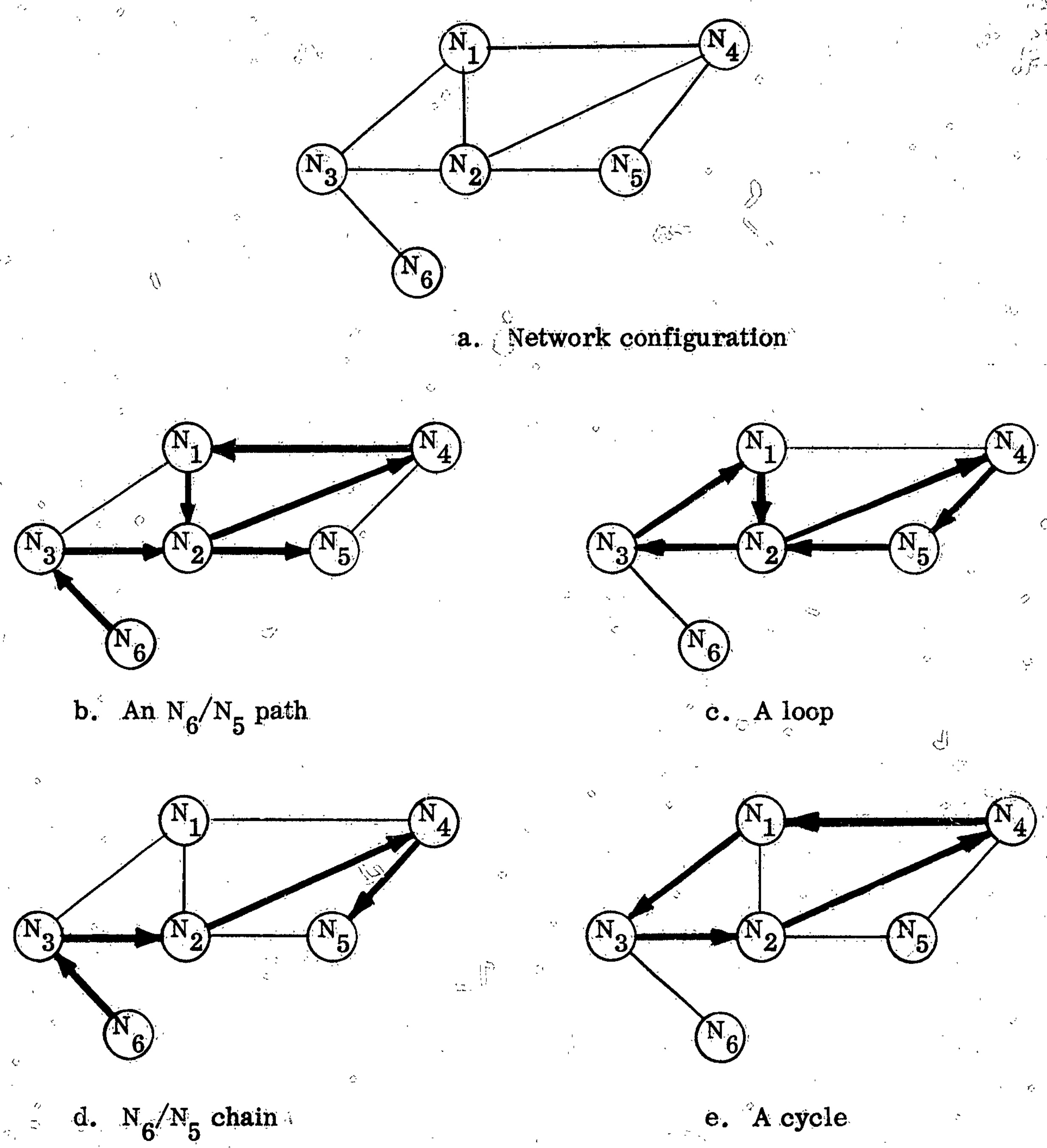
Any network can be represented by a simple geometrical structure in three-dimensional space. Let any k distinct points represent the k nodes. For each link (N_i, N_j) of the network construct a simple curve, having the corresponding points as end points, that contains none of the k-2 points corresponding to other nodes and that intersects none of the curves representing other links. Such a structure is termed a "linear graph." Figure 1a illustrates a simple network having 6 nodes and 8 links.

If A and B are distinct nodes of a network, an "A/B path" is a finite set of directed links that can be assigned an ordering $\overline{M_i, N_i}$, $i = 1, 2, \ldots, m$ (m being the number of links) so that $M_1 = A$, $N_m = B$ and $N_i = M_{i+1}$ for $i = 1, 2, \ldots, m-1$. A and B are termed the "initial" and "terminal" nodes of the path respectively. Figure 1b represents a path from N_6 to N_5 since the directed links can be arranged in the following sequence: $\overline{N_6, N_3}$, $\overline{N_3, N_2}$, $\overline{N_2, N_4}$, $\overline{N_4, N_1}$, $\overline{N_1, N_2}$, $\overline{N_2, N_5}$. A network is said to be "connected" if for every pair of distinct nodes A and B there exists at least one A/B path. Note that if a network is connected in this sense a linear graph constructed to represent the network is a connected point set in the conventional sense of connectedness. In the remainder of this paper all networks under consideration are assumed to be connected.

A "loop" in a network is a finite set of directed links that can be assigned an ordering $\overline{M_i, N_i}$, $i = 1, 2, \ldots, m$, such that $M_1 = N_m$ and $N_i = M_{i+1}$ for $i = 1, 2, \ldots, m-1$.* Figure 1c is an example of a loop.

Note that the same directed link may appear more than once in an A/B path or a loop and that both $\overline{M, N}$ and $\overline{N, M}$ may appear. If the same directed link appears k times, it is considered as being enumerated k times in the set of links that defines the path or loop.

*Note that a loop could be considered as an A/A path. In this paper it is never categorized as such, however.



a. Network configuration

b. An N_6/N_5 path

c. A loop

d. N_6/N_5 chain

e. A cycle

Fig. 1—Illustration of Basic Concepts

Since the links that form a path or loop may be arranged in a sequence of the form $\overline{M_1, M_2}$, $\overline{M_2, M_3}$, . . . , $\overline{M_{k-1}, M_k}$, it is frequently convenient to represent a path or loop by the more concise notation $\overline{M_1, M_2, \ldots, M_k}$, where the notation represents an M_1/M_k path if $M_1 \neq M_k$ and a loop if $M_1 = M_k$. It should be noted that this representation is not in general unique, however. For example $\overline{M_1, M_2, M_3, M_1, M_4, M_2, M_5}$ and $\overline{M_1, M_4, M_2, M_3, M_1, M_2, M_5}$ both represent the same M_1/M_5 path, since they are composed of the same 6 directed links.

An "A/B chain" is defined to be an A/B path such that the initial nodes (and consequently also the terminal nodes) of its directed links are all distinct. Similarly a "cycle" is a loop such that the initial (or terminal) nodes of its links are all distinct. If interpreted geometrically, chains and cycles correspond to simple or nonintersecting curves in a linear graph. Figures 1d and 1e illustrate an N_6/N_5 chain and a cycle respectively. As a result of these definitions, a chain has a unique representation of the form $\overline{M_1, M_2, \ldots, M_k}$, and the representation of a cycle is unique except for cyclic permutation of its links. (For example, $\overline{M_1, M_3, M_2, M_4, M_1}$ and $\overline{M_2, M_4, M_1, M_3, M_2}$ represent the same cycle.)

CAPACITATED AND WEIGHTED NETWORKS

If a network has k undirected links, and a nonnegative integer termed a "capacity" is associated with each of the 2k directed links, the network is termed a "capacitated network." The capacity of a directed link $\overline{M, N}$ is denoted by $c(\overline{M, N})$. The capacities need not be symmetric, i.e., it is permissible that $c(\overline{M, N}) \neq c(\overline{N, M})$. The link capacities serve as upper limits on the rate of flow through the corresponding links in the subsequent discussion. If the network is a mathematical model of a transportation network, . . . capacities might have the dimensionality "tons per unit time."

If a nonnegative real number called a "length" is associated with each of the 2k directed links of a network, the network is considered "weighted." The length of a directed link $\overline{M, N}$ is denoted by $l(\overline{M, N})$. Here again the lengths need not be symmetric. These lengths serve as weighting factors applied to the amount of flow in each link in the subsequent discussion. Depending on the application, the length of a directed link may be an abstraction of physical length, cost per unit of flow from the initial to the terminal node of the link, or some other attribute.

If α represents a path or loop in a weighted network, the "length of α", denoted by $l(\alpha)$, is defined as*

$$l(\alpha) = \sum_{\alpha} l(\overline{M, N})$$

In particular if α is the M/N path consisting of the single directed link $\overline{M, N}$, $l(\alpha) = l(\overline{M, N})$; thus there is no numerical ambiguity in using the symbol "l" for both the concepts of link length and path or loop length.

*In this equation and in other places where there is no ambiguity, the notation $\sum_{\alpha}$ is used in place of the more cumbersome notation $\sum_{\overline{M, N} \epsilon \alpha}$.

LINK FLOWS AND FLOW PATTERNS

Consider now the concepts of link flows and flow patterns in a network. A flow in a link is a nonnegative integer p associated with a particular orientation of the link. If the flow is designated as being from M to N in link (M, N), a flow f is designed for each of the directed links $\overline{M,N}$ and $\overline{N,M}$ as follows:

$$f(\overline{M,N}) = p; \quad f(\overline{N,M}) = -p$$

Thus the flow value associated with a directed link is negative if the direction of the link is counter to the assigned direction of the flow; otherwise it is positive. Note that since $f(\overline{M,N}) = -f(\overline{N,M})$ it is sufficient to prescribe the values of f for an enumerating set in order to establish flows for all directed links of the network. If $f(\overline{M,N}) = f(\overline{N,M}) = 0$, it is still convenient in some contexts to arbitrarily consider the flow as being "from" one node "to" the other.

If a flow is assigned to a link (M, N) of a capacitated network, the flow is said to be "feasible" if

$$-c(\overline{N,M}) \leq f(\overline{M,N}) \leq c(\overline{M,N})$$

or equivalent if

$$-c(\overline{M,N}) \leq f(\overline{N,M}) \leq c(\overline{N,M})$$

If $f(\overline{M,N}) = c(\overline{M,N})$, the directed link $\overline{M,N}$ is said to be "saturated" by the flow. A flow saturates both directed links $\overline{M,N}$ and $\overline{N,M}$ only in the trivial case where

$$c(\overline{M,N}) = c(\overline{N,M}) = f(\overline{M,N}) = 0*$$

*<u>For</u> if $f(\overline{M,N}) = c(\overline{M,N})$ and $f(\overline{N,M}) = c(\overline{N,M})$, it follows that $c(\overline{M,N}) + c(\overline{N,M}) = 0$, since $f(\overline{M,N}) + f(\overline{N,M}) = 0$. But since both capacities are nonnegative, both must be zero.

It is important to stress the fact that one cannot independently assign flows to each directed link of a pair, although it is evident from the preceding definitions, because $f(\overline{M,N})$ and $f(\overline{N,M})$ are two equivalent representations of the same flow, which is best thought of as a variable assigned to the undirected link (M,N).

Consider a network in which a flow is assigned to every link. Then if a node M is fixed and any enumerating set S is selected, the expression

$$\sum_{S} f(\overline{M,N}), \text{ M fixed}$$

represents the total flow away from node M minus the total flow into node M for those links whose representative in S has M as initial node. Similarly

$$\sum_{S} f(\overline{P,M}), \text{ M fixed}$$

represents the total flow into node M minus the total flow away from node M for those links whose representative in S has M as terminal node. Thus the function

$$Y(M) = \sum_{S} f(\overline{M,N}) - \sum_{S} f(\overline{P,M}), \text{ M fixed}$$

represents the total flow away from node M minus the total flow into node M. Y(M) is termed the "net output at node M." The function Y is independent of the enumerating set S since $f(\overline{M,N}) = -f(\overline{N,M})$. If the representation in S of a link is changed from $\overline{M, N}$ to $\overline{N, M}$, the effect is merely to transfer a term from one of the above summations to the other, with a corresponding change of sign.

If a flow is assigned to every link of a network and Y(P) = 0 at every node P except possibly A and B, then it is readily shown that Y(A) = -Y(B), because

$$\sum_{S} f(\overline{M, N}) - \sum_{S} f(\overline{N, M}) = 0 \tag{1}$$

since each $\overline{M, N} \epsilon S$ appears once in each summation. Also $\sum_{P \neq B} Y(P) = Y(A)$, which can be written as

$$\sum_{\substack{S \\ M \neq B}} f(\overline{M, N}) - \sum_{\substack{S \\ M \neq B}} f(\overline{N, M}) = Y(A) \tag{2}$$

Subtracting Eq. 2 from Eq. 1

$$\sum_{\substack{S \\ M = B}} f(\overline{M, N}) - \sum_{\substack{S \\ M = B}} f(\overline{N, M}) = -Y(A)$$

but the lefthand side is Y(B) by definition. If Y(A) > 0, the assignment of flows is termed an "A/B flow pattern." Y(A) is called the "value" of the flow pattern, A is termed the "source", and B is termed the "sink." If F represents an A/B flow pattern, the symbol V(F) is also used to denote its value. V(F) is necessarily an integer since all link flows are integers by definition. If Y(P) = 0 for every node P, the assignment of flows is termed a "zero flow pattern." (In some contexts it is convenient to consider a zero flow pattern as an A/B flow pattern whose value is zero.) To summarize, an assignment of flows to the links of a network is a flow pattern in this paper if, and only if, $Y(P) \neq 0$ for at most two nodes.

If $F_1, F_2, \ldots, F_n$ are A/B flow patterns (some or all of which may have the value zero) and if $f_1(\overline{M, N}), f_2(\overline{M, N}), \ldots, f_n(\overline{M, N})$ represent the corresponding flows in an arbitrary directed link $\overline{M, N}$, then the patterns are said to be "conformal" if for every link $\overline{M, N}$ either $f_i(\overline{M, N}) \geq 0$ for $i = 1, 2, \ldots, n$ or $f_i(\overline{M, N}) \leq 0$ for $i = 1, 2, \ldots, n$. Expressed differently they are conformal if there is no link such that there is nonzero flow in one direction relative to some F_i and nonzero flow in the other direction relative to some other F_j.

An algebra of flow patterns can be defined. If F and G are A/B flow patterns (where the possibility of zero flow patterns is included) and $f(\overline{M, N})$ and $g(\overline{M, N})$ denote the corresponding flows in an arbitrary directed link $\overline{M, N}$, the sum of F and G, denoted by $F \oplus G$, is defined as follows: for every directed link $\overline{M, N}$ let $h(\overline{M, N}) = f(\overline{M, N}) + g(\overline{M, N})$. The values of h then determine the link flows of $F \oplus G$. Clearly $F \oplus G$ is also an A/B flow pattern, whose value is $V(F) + V(G)$.

Similarly $F \ominus G$ is defined by the relation $h(\overline{M, N}) = f(\overline{M, N}) - g(\overline{M, N})$. In this case if $V(F) \geq V(G)$, $F \ominus G$ is an A/B flow pattern of value $V(F) - V(G)$. In the contrary case $F \ominus G$ is a B/A flow pattern of value $V(G) - V(F)$.

Elementary Flow Patterns

Certain elementary flow patterns corresponding to paths and loops play a central role in the subsequent development. If P is an A/B path or loop, if a positive flow of m units* is assigned to each link $\overline{M, N}$ of P (and necessarily also a flow of $-m$ units is assigned to $\overline{N, M}$), and if zero flow is assigned to all other links of the network, then this assignment is called an "A/B path (or loop) flow" of m units. If $m = 1$, it will also be designated a "unit path (or loop) flow." The notation m(P) or $m(\overline{M_1, M_2, \ldots, M_k})$ is used to denote such an assignment of flows. If P is an A/B path, it is readily seen that m(P) is an A/B flow pattern of value m. If P is a loop, m(P) is a zero flow pattern. Chain and cycle flows are, of course, special instances of path and loop flows respectively. The following lemma demonstrates that a unit A/B path flow can always be decomposed, in a certain sense, into simpler components.

Lemma 1: If P is an A/B path, then $1(P) = 1(C) \oplus Z$, where C is an appropriate A/B chain, Z is a zero flow pattern, and 1(C) and Z are conformal.

Proof: P can be represented as $\overline{N_1, N_2, \ldots, N_k}$, where $N_1 = A$ and $N_k = B$. If P is a chain the result follows immediately by taking $C = P$ and letting Z be the flow pattern that consists of zero flow in every link of the network. If P is not a chain, there is a first index i such that $N_i = N_j$ for some $j < i$. It is readily seen that $1(P) = 1(P_1) \oplus 1(S_1)$ where P_1 is the A/B path $\overline{N_1, \ldots, N_j, N_{i+1}, \ldots, N_k}$ and S_1 is the cycle $\overline{N_j, N_{j+1}, \ldots, N_i}$.

*If a given link $\overline{M, N}$ appears more than once, say R times, then $m \cdot R$ units of flow are to be asssigned to $\overline{M, N}$.

If P_1 is a chain, the proof is completed by setting $C = P_1$ and $Z = 1(S_1)$. If P_1 is not a chain, the above process can be repeated, expressing $1(P_1)$ as $1(P_2) \oplus 1(S_2)$ so that $1(P) = 1(P_2) \oplus 1(S_1) \oplus 1(S_2)$. So long as each successive path P_m obtained by this process is not a chain, it is always possible to "extract" another path P_{m+1} so that $1(P_m) = 1(P_{m+1}) \oplus 1(S_{m+1})$ and $1(P) = 1(P_{m+1}) \oplus 1(S_1) \oplus \ldots \oplus 1(S_{m+1})$. But clearly the process must terminate since P_{m+1} has fewer links than P_m. It follows that at some stage M, P_M is a chain. Then if $C = P_M$ and $Z = 1(S_1) \oplus \ldots \oplus 1(S_M)$, one has the desired expression for $1(P)$. Z and $1(C)$ are conformal since every link of P_M and of each of the S_m's is directed in the same sense as the corresponding link of P. This completes the proof.

Although this result is needed only in the form stated, it is true that in general $m(P) = m(C) \oplus Z'$, where C is the same A/B chain as above and Z' is obtained by multiplying all flows of Z by m.

Lemma 1 is used to establish the following general result concerning the decomposition of any A/B flow pattern.

Lemma 2: If F is an A/B flow pattern of value $k > 0$, then

$$F = 1(C_1) \oplus 1(C_2) \oplus \ldots \oplus 1(C_k) \oplus Z,$$

where the C_i's are appropriate A/B chains, Z is a zero flow pattern, and Z and the $1(C_i)$'s are conformal.

The proof involves a constructive procedure that produces the desired decomposition and makes use of the evident fact that if $Y(M) > 0$ for a node M there is at least one link $\overline{M, N}$ having positive flow, i.e., there is a link having flow directed away from M. This property is true for any assignment of flows to the links of a network, whether the assignment constitutes a flow pattern in the sense of this paper or not. (In the course of the proof, F undergoes a series of modifications that at intermediate stages determine flow assignments that are not patterns.)

Proof: Since $Y(A) = k > 0$, a link $\overline{A, N_1}$ having positive flow exists. Let F_1 denote the system of flows obtained from F by reducing the flow in $\overline{A, N_1}$ by one unit (and consequently increasing the flow in $\overline{N_1, A}$ by one unit) and leaving all other link flows unchanged. Then $Y(A) = k - 1 \geq 0$, and $Y(N_1) = 1$, relative to F_1. Hence there is a link $\overline{N_1 N_2}$ with positive flow. Let F_2 be the system of flows obtained from F_1 by reducing the flow in $\overline{N_1 N_2}$ by one unit. Then $Y(N_1) = 0$ again, but now $Y(N_2) = 1$ relative to F_2 (unless N_2 is B). So long as each successive N_j is not B, another link $\overline{N_j, N_{j+1}}$ with positive flow can be found, because either N_j is an intermediate node, in which case $Y(N_j) = 1$ relative to F_j, or else $N_j = A$. In the latter case $Y(N_j) = k$. (This is true because the preceding link was of the form $\overline{N_{j-1}, A}$, and when a unit of flow was subtracted from this link in passing from F_{j-1} to F_j, $Y(A)$ was transformed back from $k - 1$ to k.) In any case $Y(N_j) > 0$ if $N_j \neq B$. However, this procedure must terminate since at each stage a unit of flow is subtracted from some link having positive flow. Thus $N_J = B$ at some stage J. But then

$$F = F_J \oplus 1\overline{(A, N_1, \ldots, N_{J-1}, B)}$$

since the last term simply represents the collection of unit link flows removed from F to obtain F_J. Moreover F_J is an A/B flow pattern of value $k - 1$, being the difference between F and a unit A/B path flow. If $k - 1 > 0$ the above procedure can be applied to F_J, decomposing it into an A/B flow pattern of value $k - 2$ and a second unit A/B path flow. Ultimately

$$F = Z' \oplus 1(P_1) \oplus 1(P_2) \oplus \ldots \oplus 1(P_k)$$

where Z' is a zero flow pattern, the P_i's are A/B paths, and the nonzero link flows in each of these component patterns are directed the same way as in F. Applying lemma 1, each $1(P_i)$ can further be expressed as $1(C_i) \oplus Z_i$, where C_i is a chain, Z_i a zero flow pattern, and all link flows of $1(C_i)$ and Z_i are directed as in P_i. Thus

$$F = Z \oplus 1(C_1) \oplus 1(C_2) \oplus \ldots \oplus 1(C_k)$$

where $Z = Z' \oplus Z_1 \oplus Z_2 \oplus \ldots \oplus Z_k$ and the component flow patterns are all conformal. This completes the proof.

Feasible Flow Patterns

A flow pattern in a capacitated network is said to be a "feasible flow pattern" if the flow in every link is feasible, i.e., if

$$-c(\overline{N, M}) \leq f(\overline{M, N}) \leq c(\overline{M, N})$$

for every directed link $\overline{M, N}$. There are only a finite number of distinct feasible flow patterns that can be associated with a given capacitated network. Letting $C = \max c(\overline{M, N})$ and L be the number of undirected links, there are at most

$2C + 1$ distinct flows possible in any given link, viz., flows of $1, 2, \ldots, C$ units in either direction or zero flow. Thus there are at most $(2C + 1)^L$ ways to assign flows to all links.* Of course most of these assignments are not flow patterns in the sense of this paper.

Lemma 3, which establishes a useful result concerning feasible flow patterns states:

Lemma 3: If F is a feasible A/B flow pattern and $G_1, G_2, \ldots, G_k$ are conformal A/B flow patterns such that $F \oplus G_1 \oplus G_2 \oplus \ldots \oplus G_k$ is feasible, then $F \oplus S$ is also a feasible A/B flow pattern if S is the sum of any subset of the G_i's.

Proof: Let $f, g_1, g_2, \ldots, g_k$ and s denote the corresponding flows in an arbitrary directed link $\overline{M, N}$. Since the g_i's are all nonnegative or all nonpositive, one of the following relations is true:

$$f \leq f + s \leq f + g_1 + g_2 + \ldots + g_k$$

or

$$f + g_1 + g_2 + \ldots + g_k \leq f + s \leq f$$

*More precisely there are exactly $\Pi [c(\overline{M, N}) + c(\overline{N, M}) + 1]$ ways to assign flows to all links of a network, where the product ranges over all undirected links (M, N).

The feasibility assumptions of the lemma also imply that both the following relations hold:

$$-c(\overline{N, M}) \leq f \leq c(\overline{M, N})$$

and

$$-c(\overline{N, M}) \leq f + g_1 + g_2 + \ldots + g_k \leq c(\overline{M, N})$$

It follows that $-c(\overline{N, M}) \leq f + s \leq c(\overline{M, N})$; i.e., the flow in $\overline{M, N}$ corresponding to the flow pattern $F \oplus S$ is feasible. But since $\overline{M, N}$ was arbitrary it follows that this flow pattern is feasible. This completes the proof.

Using lemmas 2 and 3 the self-evident fact can be rigorously established that if a feasible A/B flow pattern of value k exists in a capacitated network, feasible A/B flow patterns having values 0, 1, 2, . . . , k exist. For if F_k is feasible and $V(F_k) = k$, F_k can be expressed as $Z \oplus 1(C_1) \oplus \ldots \oplus 1(C_k)$ according to lemma 2. If F_i is defined as $Z \oplus 1(C_1) \oplus \ldots \oplus 1(C_i)$ for $i = 1, 2, \ldots, k$, and $F_0 = Z$, according to lemma 3 F_i is feasible for $i = 0, 1, \ldots, k$. Moreover, since Z and the $1(C_i)$'s are conformal, the F_i's are nested flow patterns in the sense that the magnitude of flow in each undirected link is monotonic nondecreasing and its direction remains the same, as i increases from 0 to k.

A "maximal A/B flow pattern" is a feasible A/B flow pattern F whose value V(F) is a maximum. Since there are only a finite number of feasible flow patterns possible in a capacitated network, a maximal A/B flow pattern exists for every pair of distinct nodes A and B. (Note that the flow pattern that consists of zero flow in every link is necessarily feasible and can be considered an A/B flow pattern for any distinct nodes A and B.) The value of a maximal A/B flow pattern is called the "A/B capacity of the network."

"Cost" of a Flow Pattern

Consider next an A/B flow pattern F in a weighted network. F is associated with the nonnegative integer T(F) defined by the relation:

$$T(F) = \sum_{S} [|f(\overline{M,N})| \cdot l'(M,N)]$$

where S is an arbitrary enumerating set and $l'(M, N)$ is defined as

$$l'(M,N) = l(\overline{M,N}) \text{ if } f(M,N) \geq 0$$

and

$$l'(M,N) = l(\overline{N,M}) \text{ if } f(M,N) < 0$$

Thus T(F) is obtained by multiplying the flow in each link by the length, measured in the direction of the flow, and summing over all links of the network. If $l(\overline{M,N})$ is interpreted as the cost of moving one unit of flow from M to N through $\overline{M,N}$, T(F) is simply the total cost associated with the flow pattern F. T(F) is necessarily nonnegative.

The following lemma expresses an important relation between the costs of certain flow patterns.

Lemma 4: If F, G, and H are A/B flow patterns and G and H are conformal, then

$$T(F \oplus G \oplus H) - T(F \oplus G) \geq T(F \oplus H) - T(F)$$

Proof: Let $Q = \tilde{T}(F \oplus G \oplus H) - T(F \oplus G) - T(F \oplus H) + T(F)$. It is desired to prove that $Q \geq 0$ under the assumptions of this lemma. Let S be an enumerating set chosen in such a way that $f(\overline{C, D}) \geq 0$ for all $\overline{C, D} \epsilon S$, where $f(\overline{C, D})$ denotes the flow in $\overline{C, D}$ corresponding to pattern F. In the following discussion an arbitrary $\overline{C, D} \epsilon S$ is fixed. For conciseness the flows in $\overline{C, D}$ corresponding to the three patterns are denoted by f, g, and h and the lengths $l(\overline{C, D})$ and $l(\overline{D, C})$ by l and l' respectively. If $Q(\overline{C, D})$ is defined by the relation

$$Q(\overline{C, D}) = |f + g + h| x_1 - |f + g| x_2 - |f + h| x_3 + |f| x_4$$

where each x_i is l or l', depending on whether the corresponding expression within absolute value signs is positive or negative, $Q(\overline{C, D})$ represents the contribution to Q corresponding to $\overline{C, D}$ and $\sum_{\overline{C, D} \epsilon S} Q(\overline{C, D}) = Q$. Thus it is sufficient to show that each $Q(\overline{C, D}) \geq 0$.

One of the following sets of relations holds:

$$f \geq 0, \; g \geq 0, \; h \geq 0 \tag{3}$$

or

$$f \geq 0, \; g \leq 0, \; h \leq 0 \tag{4}$$

since G and H are conformal.

If relation 3 holds, $Q(\overline{C, D}) = (f + g + h)\, l - (f + g)\, l - (f + h)\, l + f \cdot l = 0$. If relation 4 holds and if $f + g + h \geq 0$, then $f + g \geq 0$ and $f + h \geq 0$ also, and $Q(\overline{C, D})$ reduces to the same expression as for relation 3, having the value zero.

It remains to investigate the case when relation 4 holds and $f + g + h < 0$. In this case $Q(\overline{C, D})$ reduces to one of the following expressions, depending on the relative magnitudes of f, g, and h:

$$-(f + g + h)l' - (f + g)l - (f + h)l + f \cdot l$$

$$-(f + g + h)l' + (f + g)l' - (f + h)l + f \cdot l$$

$$-(f + g + h)l' - (f + g)l + (f + h)l' + f \cdot l$$

$$-(f + g + h)l' + (f + g)l' + (f + h)l' + f \cdot l$$

When simplified these reduce to the expressions

$$-(f + g + h)(l' + l)$$

$$-h(l' + l)$$

$$-g(l' + l)$$

$$f(l' + l)$$

These are all nonnegative since $f + g + h < 0$, $l' + l \geq 0$, and the conditions of relation 4 hold. Thus $Q(\overline{C, D}) \geq 0$ in all cases. This completes the proof.

Thus the increment of cost resulting from adding H to $F \oplus G$ is at least as great as that resulting from adding H to F alone. (This result fails to hold if G and H are not conformal.)

Ideal Flow Patterns

Finally consider a network that is both capacitated and weighted, and let A and B be two fixed, distinct nodes of the network. If a feasible A/B flow pattern of value k exists, then one, call it F, exists such that T(F) is minimum, since the feasible A/B patterns of value k are finite in number. Such a pattern is termed an "ideal A/B flow pattern of value k." If the network is interpreted as a mathematical model of a transportation network, an ideal A/B flow pattern of value k can be interpreted as a pattern that moves k units of flow per unit time from A to B at minimum cost. Thus ideal flow patterns are in this sense "best" flow patterns.

The following lemma establishes an importation relation between certain ideal flow patterns.

Lemma 5: If F_{i-1}, F_i, and F_{i+1} are ideal flow patterns whose values are i-1, i, and i+1 respectively, then

$$T(F_{i+1}) - T(F_i) \geq T(F_i) - T(F_{i-1})$$

Proof: $F_{i+1} \ominus F_{i-1}$ is an A/B flow pattern of value 2. As such it can (according to lemma 2) be expressed as $1(C_1) \oplus 1(C_2) \oplus Z$ where C_1 and C_2 are A/B chains; Z is a zero flow pattern; and $1(C_1)$, $1(C_2)$, and Z are conformal. Since F_{i-1} and $[F_{i-1} \oplus 1(C_1) \oplus 1(C_2) \oplus Z]$ are feasible patterns (the latter being simply F_{i+1}), it follows (according to lemma 4) that $T(F_{i+1}) \geq T[F_{i-1} \oplus 1(C_1) \oplus 1(C_2)] + T(F_{i-1} \oplus Z) - T(F_{i-1})$. But $F_{i-1} \oplus Z$ is an A/B flow pattern of value i-1 and is feasible (according to lemma 3). Hence $T(F_{i-1} \oplus Z) - T(F_{i-1}) \geq 0$ since F_{i-1} is ideal. It follows that

$$T(F_{i+1}) \geq T[F_{i-1} \oplus 1(C_1) \oplus 1(C_2)]$$

Applying lemma 4 again

$$T[F_{i-1} \oplus 1(C_1) \oplus 1(C_2)] \geq T[F_{i-1} \oplus 1(C_1)] + T[F_{i-1} \oplus 1(C_2)] - T(F_{i-1})$$

But $F_{i-1} \oplus 1(C_1)$ and $F_{i-1} \oplus 1(C_2)$ are A/B flow patterns of value i and are feasible (according to lemma 3). Since F_i is ideal, $T[F_{i-1} \oplus 1(C_1)] \geq T(F_i)$ and $T[F_{i-1} \oplus 1(C_2)] \geq T(F_i)$. Hence

$$T(F_{i+1}) \geq 2T(F_i) - T(F_{i-1})$$

But this is equivalent to the inequality that was to be established.

A sequence of real numbers $a_0, a_1, a_2, \ldots, a_k$ is said to be convex if

$$a_{i+1} - a_i \geq a_i - a_{i-1}$$

for $i = 1, 2, \ldots, k-1$. If $F_0, F_1, F_2, \ldots, F_k$ are ideal A/B flow patterns of values $0, 1, 2, \ldots, k$ respectively, where k is the A/B capacity of the network, it follows that the sequence $T(F_0), T(F_1), \ldots, T(F_k)$ is a convex sequence of nonnegative real numbers, the first of which is zero. [The flow pattern F_0, which assigns zero flow to every link, is clearly ideal, and $T(F_0) = 0$.]

In a certain sense the convex sequence $T(F_0), T(F_1), \ldots, T(F_k)$ can be considered as the "A/B cost profile" for a given network. The remainder of this paper is devoted to developing a general method for determining the A/B cost profile for a given network and given nodes A and B, and for finding specific ideal flow patterns possessing these minimum costs. The problem is precisely defined in the next section.

As a matter of interest, note that given any finite convex sequence of nonnegative real numbers $a_0, a_1, a_2, \ldots, a_n$ a network exists having this sequence as its A/B cost profile. In fact the network of Fig. 2 has the desired properties. In this figure the numbers indicate the lengths, in both directions, of the corresponding links. All link capacities are assumed to be one unit. If $1 \leq i \leq n$, the (in this case unique) ideal A/B flow pattern of value i consists of assigning one unit of flow from A to B in each of the paths $\overline{A, N_j, B}$, $j = 1, 2, \ldots, i$, and assigning zero flow in the remaining links. Clearly $T(F) = a_i$ for this pattern. (The intermediate nodes $N_1, N_2, \ldots, N_n$ are introduced merely to conform with the requirement that at most one link join two distinct nodes.)

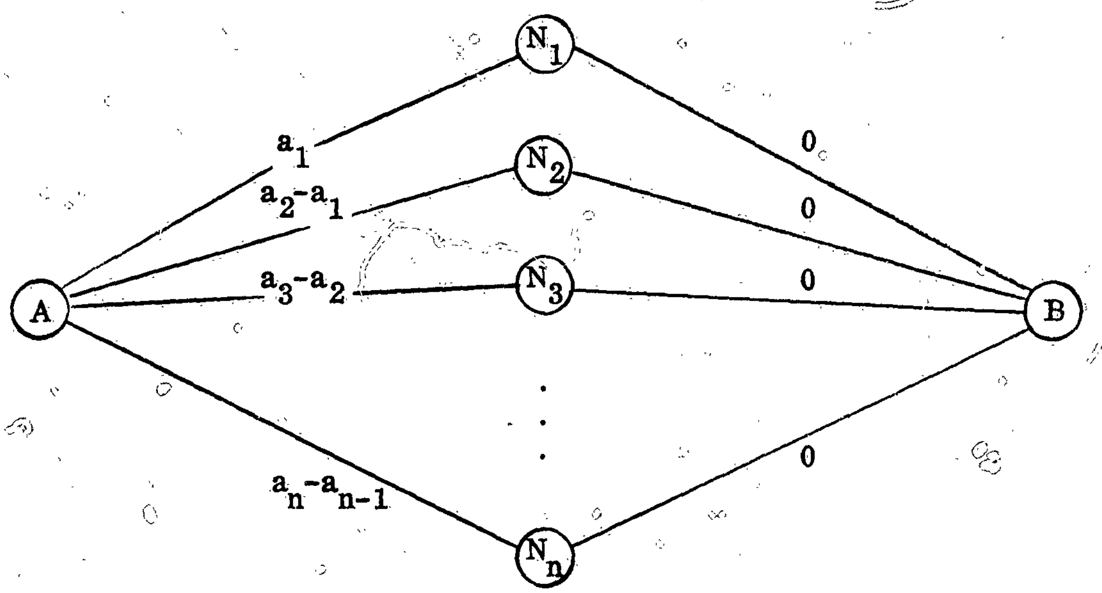


Fig. 2--A Network Having A Given A/B Cost Profile

STATEMENT OF PROBLEM AND SOLUTION IN PRINCIPLE

FORMAL STATEMENT OF PROBLEM

In the remainder of this paper the network under consideration is assumed to be connected, weighted, and capacitated and to have the property that for every cycle S, $l(S) > 0$.

The problem of immediate concern can be precisely stated:

> Given a network of the above type, and two distinguished nodes A and B, produce a procedure that will
>
> (a) determine the A/B capacity k of the network;
>
> (b) determine a specific ideal A/B flow pattern F_i of value i, for $i = 1, 2, \ldots, k$; and
>
> (c) always terminate in a finite number of steps.

Network flow problems of various types have been the subject of considerable study and a number of procedures having property (c), that can be used to solve problem (a), or (b) for any specified value of the flow pattern, exist in the literature. Reference is made to some of these after the present procedure is defined and its validity established.

In this section a procedure solving the above problem in principle is given, accompanied by the proofs required to establish its validity. The development is incomplete in one respect, however. The assumption is made that it is possible to design an algorithm for finding A/B chains of a certain type. The next section of the paper produces this algorithm, which completes the development.

AN EXAMPLE

The following example motivates the central concept involved in the procedure, viz., the effective length of a chain.

Consider the network shown in Fig. 3a. Here A, N_2, N_3, N_4, and B represent the nodes, and there are seven links as indicated. The numbers above and below each link represent its length and capacity respectively. (These apply to both directed links; i.e., the network is fully symmetric.) An ideal A/B flow pattern of value 0 is of course the pattern F_0, which assigns zero flow to every link. The unit A/B chain flow $F_1 = 1(\overline{A, N_2, N_3, N_4, B})$ is an ideal pattern of value 1. It happens to be the only one for this network. Note that $T(F_1) = 4$ and that F_1 saturates the directed links $\overline{A, N_2}$, $\overline{N_2, N_3}$, and $\overline{N_3, N_4}$.

One possible way to augment F_1 and obtain an A/B pattern of value 2, G_2, would be to add $1(\overline{A, N_4, B})$ to F_1. The resulting pattern is shown in Fig. 3b, where the numbers and arrows indicate the magnitudes and directions of link flows. But G_2 is not ideal since $T(G_2) = 11$, whereas $T(F_2) = 10$, where F_2 is the feasible A/B flow pattern shown in Fig. 3c. F_2 is ideal, and there is no other ideal pattern of value 2 in this network. F_2 is also maximal since clearly link $\overline{N_4, B}$ limits flow to two units at most. Thus the cost profile for this network is 0, 4, and 10.

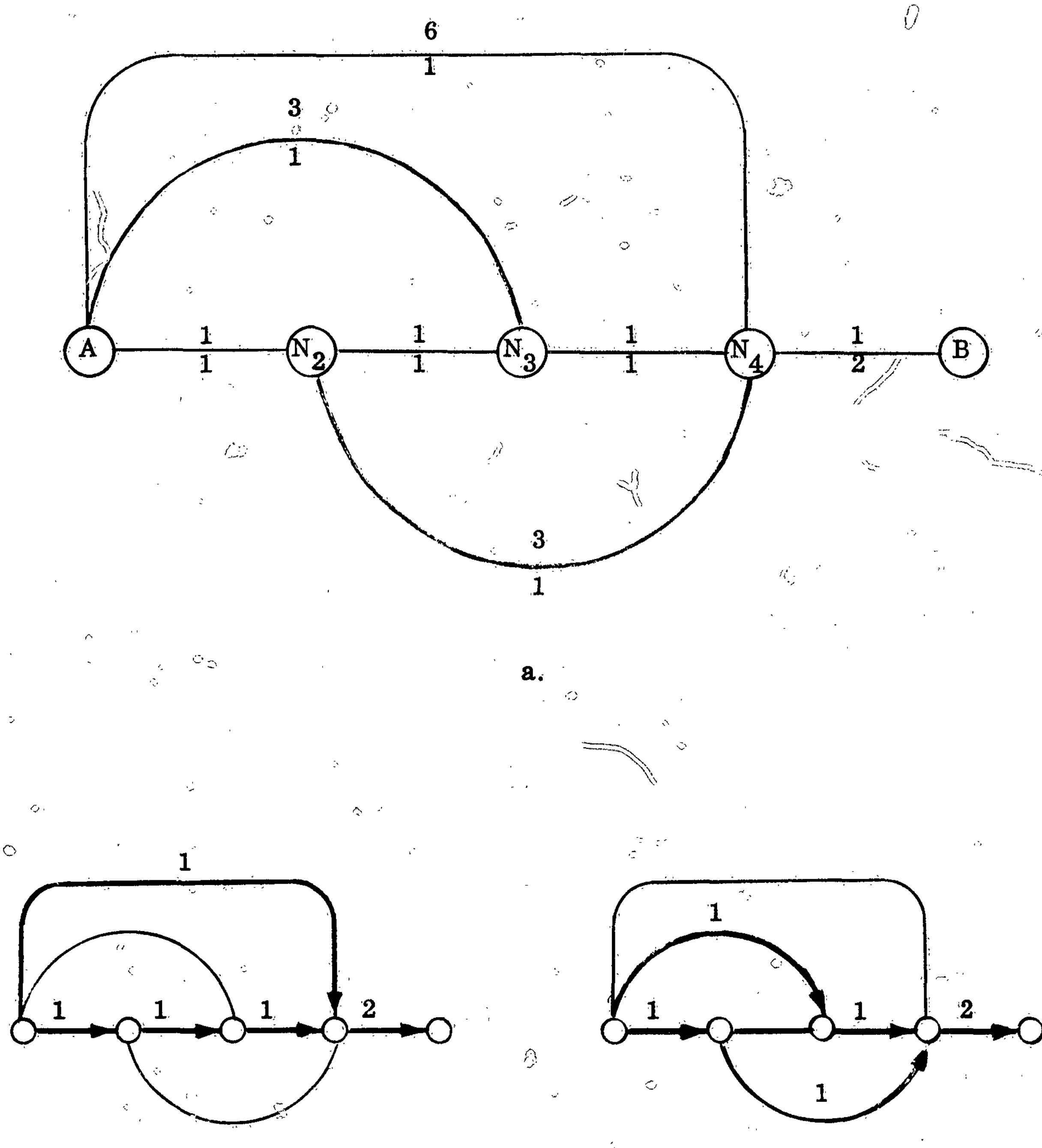


Fig. 3--Comparable A/B Flow Patterns

Now considering the statement of the problem, especially property (b), it is highly desirable that the procedure developed involves using F_i, once it is found, as the starting point for constructing F_{i+1} rather than constructing F_{i+1} independently.

If F_2 is to be obtained by augmenting F_1 in the example, the pattern that must be added to F_1 is $F_2 \ominus F_1 = 1(\overline{A, N_3, N_2, N_4, B})$. It is not coincidental that $F_2 \ominus F_1$ is a unit A/B chain flow. It is shown below that if F_i is ideal and non-maximal, an ideal pattern F_{i+1} can always be obtained by adding an appropriate unit A/B chain flow. There may be other ways also, of course. In the case of the present network, since F_1 and F_2 are unique, it follows that $F_2 \ominus F_1$ had to be a unit A/B chain flow.

Once the above result is rigorously established, viz., that it is sufficient to consider patterns of the form $F_i \oplus 1(C)$ in generating F_{i+1}, it remains to establish a simple criterion for determining which chain to use and to develop a systematic method for finding a chain satisfying the criterion.

Consider again the chains involved in generating G_2 and F_2 from F_1 in the example. These are $C_1 = \overline{A, N_4, B}$ and $C_2 = \overline{A, N_3, N_2, N_4, B}$ respectively. In terms of length, C_1 is preferable since $l(C_1) = 7$ whereas $l(C_2) = 8$. In terms of increments of T, however, $T(G_2) - T(F_1) = 7$ and $T(F_2) - T(F_1) = 6$. Thus C_2 is <u>effectively</u> shorter. This notion is now made general and precise.

EFFECTIVE LENGTH

If F is a feasible A/B flow pattern with associated link flows denoted by $f(\overline{M, N})$, where $\overline{M, N}$ is an arbitrary directed link, the "effective length of

$\overline{M, N}$ relative to F", denoted by $e(\overline{M, N}; F)$, is defined as:

$$e(\overline{M, N}; F) \begin{cases} = \ 1(\overline{M, N}) \text{ if } 0 \leq f(\overline{M, N}) < c(\overline{M, N}) \\ = \ -1(\overline{N, M}) \text{ if } f(\overline{M, N}) < 0 \\ = \ \infty^* \text{ if } f(\overline{M, N}) = c(\overline{M, N}) \end{cases}$$

If C denotes any chain or cycle, the "effective length of C relative to F", denoted by $e(C;F)$, is defined as:

$$e(C;F) = \sum_{C} e(\overline{M, N}; F)$$

Note that $e(\overline{M, N}; F) = \infty$ if and only if F saturates $\overline{M, N}$, and consequently $e(C;F) < \infty$ if and only if no link of C is saturated. This leads to the following result:

Lemma 6: If F is a feasible A/B flow pattern and C is an A/B chain or a cycle, $F \oplus 1(C)$ is feasible if and only if $e(C;F)$ is finite.

Proof: A link $\overline{M, N}$ is saturated if and only if $e(\overline{M, N}; F) = \infty$. Thus if $e(C;F)$ is finite, no link of C is saturated, and so the flow in each link of C can be increased by at least one unit. Hence $F \oplus 1(C)$ is feasible if $e(C;F)$ is finite. Conversely if $F \oplus 1(C)$ is feasible no link of C is saturated, and $e(C;F)$ is finite since $e(\overline{M, N}; F)$ is finite for $\overline{M, N} \in C$. This completes the proof.

*For convenience "∞" is adjoined to the real number system. The only properties of this element that are required are (a) if R is any real number, $R < \infty$; (b) it is not true that $\infty < \infty$; (c) $R + \infty = \infty$; (d) $\infty + \infty = \infty$; and (e) $\min\{\infty, \infty, \ldots, \infty\} = \infty$.

Lemma 7 is a formal statement of the fact that the definition of effective length satisfies the property suggesting its introduction.

Lemma 7: If F is a feasible A/B flow pattern and C is an A/B chain or a cycle such that $e(C;F)$ is finite, then $T[F \oplus 1(C)] - T(F) = e(C;F)$.

Proof: Consider any $\overline{M,N} \in C$. If $f(\overline{M,N}) \geq 0$, where f denotes flow corresponding to F, then the flow in $\overline{M,N}$ corresponding to $F \oplus 1(C)$ is $f(\overline{M,N}) + 1$ and T is increased in this link by the amount $l(\overline{M,N})$. If $f(\overline{M,N}) < 0$, the flow in $\overline{M,N}$ corresponding to $F \oplus 1(C)$ is $f(\overline{M,N}) + 1 \leq 0$ and that in $\overline{N,M}$ is $f(\overline{N,M}) - 1 \geq 0$; thus T is decreased in this link by $l(\overline{N,M})$ units or increased by $-l(\overline{N,M})$. In either case the increase in T with respect to $\overline{M,N}$ equals $e(\overline{M,N};F)$ as defined above. Since $\overline{M,N}$ was arbitrary, the result follows.

The following result, which is an immediate consequence of lemmas 6 and 7, plays an important part in the next section.

Lemma 8: If F is an ideal A/B flow pattern, then $e(S;F) \geq 0$ for every cycle S.

Proof: According to lemma 7 $T[F \oplus 1(S)] = T(F) + e(S;F)$. If $e(S;F) < 0$ it follows that $T[F \oplus 1(S)] < T(F)$. But $F \oplus 1(S)$ is a feasible A/B flow pattern according to lemma 6 and has value V(F), which contradicts the assumption that F is ideal.

CENTRAL THEOREMS

The set of maximal A/B flow patterns can be characterized in terms of the notion of effective length in the following way.

Theorem 1: A feasible A/B flow pattern F is maximal if and only if $e(C;F) = \infty$ for every A/B chain C.

Proof: If $e(C;F)$ is finite for some A/B chain C, then $F \oplus 1(C)$ is feasible by lemma 6, and thus F is not maximal. Hence if F is maximal, $e(C;F) = \infty$ for all A/B chains. Conversely if F is not maximal, a feasible A/B flow pattern G of value $V(F) + 1$ exists. According to lemma 2 $G \ominus F$ can be written as $1(C) \oplus Z$, where C is an A/B chain, Z is a zero flow pattern, and $1(C)$ and Z are conformal. Since F and $F \oplus 1(C) \oplus Z = G$ are feasible patterns, it follows from lemma 3 that $F \oplus 1(C)$ is feasible also. Thus $e(C;F)$ is finite according to lemma 6. Hence if $e(C;F) = \infty$ for all A/B chains, F is maximal. This completes the proof.

Since there are only a finite number of distinct A/B chains in a network and each has a well-defined effective length relative to any given feasible A/B flow pattern F, a chain C_1 exists that minimizes $e(C;F)$. This chain is referred to as a "chain of shortest effective length relative to F", or as an "effectively shortest" chain. It is a consequence of theorem 1 that $e(C_1;F) < \infty$ if and only if F is nonmaximal.

An ideal A/B flow pattern that is nonmaximal can always be extended to an ideal pattern of value one unit greater in a manner described in the next theorem.

Theorem 2: If F is an ideal nonmaximal A/B flow pattern of value k and C is any A/B chain of shortest effective length relative to F, then $F \oplus 1(C)$ is an ideal A/B flow pattern of value k + 1.

Proof: Let G be an ideal flow pattern of value k + 1. Then according to lemma 2 $G \ominus F$ can be expressed as $1(C') \oplus Z$, where C' is an A/B chain, Z is a zero flow pattern, and 1(C') and Z are conformal. Since F and $F \oplus 1(C') \oplus Z$ are feasible, lemma 4 asserts that the following inequality holds:

$$T(G) - T[F \oplus 1(C')] \geq T(F \oplus Z) - T(F)$$

But $F \oplus Z$ is feasible, according to lemma 3, so that $T(F \oplus Z) - T(F) \geq 0$ since F is ideal. Thus $T(G) \geq T[F \oplus 1(C')]$. Now $F \oplus 1(C')$ is also feasible, according to lemma 3, and has value k + 1. It follows that $F \oplus 1(C')$ is also ideal. C' must minimize e(C;F) because if e(C";F) < e(C';F) it would follow from lemma 7 that $T[F \oplus 1(C'')] < T[F \oplus 1(C')]$. It is also clear from lemma 7 that if C is any other chain with e(C;F) = e(C';F), then $F \oplus 1(C)$ is an ideal A/B flow pattern whose value is k + 1. This completes the proof.

If the existence of a terminating algorithm that will always find a chain of shortest effective length relative to an ideal flow pattern F, when F is non-maximal, and will give a suitable indication when F is maximal is assumed, then theorems 1 and 2 could be used to form the basis for a procedure for solving the stated problem. However, this would require proceeding in increments of one unit of flow—a severe limitation in any practical problem.

Fortunately the convexity property established in lemma 5 can be used to formulate a general method for accelerating the process.

Let F_i be an ideal A/B flow pattern of value i in a network whose A/B capacity is at least i + 2. Then if the A/B chain C_1 minimizes $e(C;F)$, $F_{i+1} = F_i \oplus 1(C_1)$ is ideal according to theorem 2. Similarly $F_{i+2} = F_{i+1} \oplus 1(C_2)$ is ideal if C_2 minimizes $e(C;F_{i+1})$. Applying lemma 7 the following equations hold:

$$T(F_{i+2}) - T(F_{i+1}) = e(C_2;F_{i+1})$$

and

$$T(F_{i+1}) - T(F_i) = e(C_1;F_i)$$

But lemma 5 asserts that

$$T(F_{i+2}) - T(F_{i+1}) \geq T(F_{i+1}) - T(F_i)$$

It follows that

$$e(C_2;F_{i+1}) \geq e(C_1;F_i)$$

Now suppose that $e(C_1;F_{i+1}) = e(C_1;F_i)$. Since C_2 minimizes $e(C;F_{i+1})$, it follows that C_1 is an effectively shortest chain relative to F_{i+1} also, so that $G_{i+2} = F_{i+1} \oplus 1(C_1) = F_i \oplus 2(C_1)$ is an ideal A/B pattern of value i + 2. In general once a chain C_1 that minimizes $e(C;F_i)$ has been found ideal patterns F_{i+1}, F_{i+2}, . . . can be generated by adding $1(C_1)$ to each preceding pattern. As long as $e(C_1;F_{i+k-1}) = e(C_1;F_i)$, it will be true that $F_{i+k} = F_{i+k-1} \oplus 1(C_1)$ $= F_i \oplus k(C_1)$ is an ideal A/B pattern of value i + k. Theorem 3 establishes a criterion for determining the maximum number of times that the same unit chain flow $1(C_1)$ can be added to an ideal flow pattern before the resulting pattern ceases to be ideal.

Theorem 3: If F is an ideal nonmaximal A/B flow pattern and C is any A/B chain of shortest effective length relative to F, then $F \oplus j(C)$ is an ideal A/B flow pattern of value $V(F) + j$, for $j = 1, 2, \ldots, Q$, where $Q = \min_{C} q(\overline{M,N};F)$ and $q(\overline{M,N};F)$ is defined as

$$q(\overline{M,N};F) = \begin{array}{l} c(\overline{M,N}) - f(\overline{M,N}) \text{ if } f(\overline{M,N}) \geq 0 \\ -f(\overline{M,N}) \text{ if } f(\overline{M,N}) < 0 \end{array}$$

Proof: From theorem 1 it is known that $e(\overline{M,N};F)$ is finite for every $\overline{M,N} \epsilon C$. Referring to the definition of $e(\overline{M,N};F)$, two cases must be considered.

(a) $f(\overline{M,N}) \geq 0$. In this case $e(\overline{M,N};F) = l(\overline{M,N})$. In fact $e[\overline{M,N};F \oplus k(C)] = l(\overline{M,N})$ so long as $f(\overline{M,N}) + k < c(\overline{M,N})$; i.e., so long as $k < c(\overline{M,N}) - f(\overline{M,N})$. $1(C)$ can be added to F $c(\overline{M,N}) - f(\overline{M,N}) - 1$ times without altering the effective length of $\overline{M,N}$. Thus $1(C)$ can be added to F one more time, [i.e., a total of $q(\overline{M,N};F)$ times] before the effective length changes. In this case $\overline{M,N}$ is saturated and $e[\overline{M,N};F \oplus q(C)] = \infty$.

(b) $f(\overline{M,N}) < 0$. In this case $e(\overline{M,N};F) = -l(\overline{N,M})$. In fact $e[\overline{M,N};F \oplus k(C)] = -l(\overline{N,M})$ so long as $k < -f(\overline{M,N})$. Thus the effective length of $\overline{M,N}$ changes only after adding $1(C)$ to F, $-f(\overline{M,N})$ times, which is the same as $q(\overline{M,N};F)$ times.

In this case the flow in (M,N) is then reduced to zero, and the effective length of $\overline{M,N}$ changes from $-l(\overline{N,M})$ to $l(\overline{M,N})$.

Now if Q is taken to be the minimum of the link q's, it follows that for $j = 1, 2, \ldots, Q-1$

$$e[C;F \oplus j(C)] = e(C;F)$$

since this relation is true for every directed link of C. But it follows from the remarks immediately preceding the statement of theorem 3 that so long as the effective length of C remains constant the addition of 1(C) to each successive pattern generates another ideal pattern. This completes the proof.

PROCEDURE

The existence of an algorithm, which is referred to as the "effective length algorithm", is assumed. As input it requires an enumerating set S characterizing the network configuration and an ideal A/B flow pattern F_i of value i, described by itemizing the flow $f_i(\overline{M,N})$ for all $\overline{M,N} \in S$. As output it produces the effective distance D_i from A to B relative to F_i.* If $D_i < \infty$ (i.e., if F_i is nonmaximal) it also produces a specific effectively shortest chain C_i.

A procedure that will always solve the stated problem can now be formally presented. The procedure involves repeating the following steps as

*"Effective distance" means the effective length of an effectively shortest chain.

many times as required to reach a maximal flow pattern. At the start take $i = 0$ and take as F_0 the ideal A/B flow pattern for which $f_0(\overline{M, N}) = 0$ for all $\overline{M, N} \epsilon S$.

(a) Apply the effective distance algorithm to the network with flow pattern F_i. If $D_i = \infty$ (i.e., if F_i is maximal), then terminate the procedure because the complete solution to the problem is obtained. Otherwise, perform step b.

(b) Let C_i be the effectively shortest chain found in a. Scan the links of C_i and determine $Q_i = \min [q(\overline{M, N}); F_i]$ as defined in theorem 3.

(c) Construct an ideal A/B flow pattern F_{i+k} of value $i + k$ as follows for $k = 1, 2, \ldots, Q_i$:

For every $\overline{M, N} \epsilon S$ define f_{i+k} as

(1) $f_i + k$ if $\overline{M, N} \epsilon C_i$,

(2) $f_i - k$ if $\overline{N, M} \epsilon C_i$,

(3) f_i otherwise

Then repeat (a) with $i + Q_i$ replacing i and F_{i+Q_i} replacing F_i.

Theorem 3 states that the flow patterns obtained in this way are ideal, and according to theorem 1 the process will terminate when and only when, after step c is completed, the resulting F_{i+Q_i} is maximal.

RELATION TO OTHER PROCEDURES

Maximizing Flow. The principle of obtaining a maximal flow pattern by adding a succession of unsaturated A/B chains is well known. Dantzig and

Fulkerson[1] present an efficient hand-computing scheme using this principle, which is applicable to any capacitated network with symmetric capacities [except that $c(M, N) = 0$ if $M = B$ or $N = A$]. Any procedure for determining unsaturated A/B chains could be made the basis for a flow-maximizing process. The present procedure is a special instance of this approach since it selects from all A/B chains C one that minimizes $e(C;F)$.

Minimizing Cost for Stated Flow. If k is a nonnegative integer that does not exceed the A/B capacity of the network, the problem of finding a feasible A/B flow pattern F such that $V(F) = k$ and $T(F)$ is minimum is a special case of the general linear-programing problem. Thus, in theory, a technique such as the simplex method could be employed. However, in complex networks the number of variables and constraints is so large as to make such a general approach infeasible.

This problem is also a special case of the class of network-flow problems known as "capacity-constrained transshipment problems." Fulkerson[2] describes this type of problem and shows that it is equivalent to an appropriate Hitchcock problem. Thus any procedure for solving the Hitchcock problem could also be employed.

Finding a Complete Set of Ideal Flows. The problem of finding a family of ideal A/B flow patterns, one for every feasible value of flow, is related to a dynamic problem posed and solved by Ford and Fulkerson of the RAND Corporation.[3] That problem deals with maximizing the total flow arriving at a node B by the end of T time periods assuming there is no flow in transit at time $T = 0$. (The lengths of links are assumed to be positive integers and are interpreted as the number of time periods required to traverse the links.) The procedure presented finds a family of solutions—one for every value of T. (These solutions remain the same after a certain value of T is reached.) The procedure involves finding static flow patterns in the network, and it is shown that these flow patterns are ideal, at least when T is sufficiently large.

<u>Decomposing Solutions into Routes</u>. Although the procedure described in this paper does not include a specific method for expressing the ideal flow patterns as the sum of conformal chain flows,* it is recognized that this is desirable if one wants an operational plan for routing shipments at minimum cost. The construction employed in the proof of lemma 2 could be formalized into an algorithm for producing such a composition.† However, efficient schemes for decomposing flows exist. Specifically the labeling process termed "Routine II" by Ford and Fulkerson[3] performs this function efficiently.

*The example given at the beginning of the present section shows that the chain flows used to synthesize an ideal pattern are not necessarily conformal.

†Referring to lemma 2 the residual flow pattern Z is necessarily the pattern that is identically zero. This can be shown to be a consequence of the assumption that every cycle C has positive length. If Z were not identically zero, the pattern F being decomposed could not be ideal.

STATEMENT OF THE ALGORITHM

Let F be an ideal A/B flow pattern in a network, with link flows denoted by $f(\overline{M,N})$. Consider the links of the network to be arranged in a sequence (M_i,N_i), $i = 1, 2, \ldots, L$. The algorithm consists of the assignment and revision of certain quantities, or "labels", associated with the nodes. It is patterned after the labeling procedures employed by Ford and Fulkerson for solving various network-flow problems. Ford[4] describes a labeling process concerned with finding an A/B chain C that minimizes $l(C) = \sum_C l(\overline{M,N})$. The present algorithm is an adaptation of that one, where the necessary modifications were made to minimize $\sum_C e(\overline{M,N};F)$ rather than $\sum_C l(\overline{M,N})$. In addition information is carried along that enables one to readily determine a specific chain that minimizes $\sum_C e(\overline{M,N};F)$. This is achieved by recording certain approach links in the same manner as the labeling process of Ford and Fulkerson.[5] Formal description of the effective-length algorithm follows.

(a) Associate with node A the label $D(A) = 0$, and let $D(N) = \infty$ for all other nodes.

(b) Considering each link (M_i,N_i) in turn, perform the following adjustments:

(1) If $D(N_i) > D(M_i) + e(\overline{M_i,N_i};F)$, replace $D(N_i)$ by this smaller quantity and record $\overline{M_i,N_i}$ as the approach link currently associated with node N_i.

(2) If $D(M_i) > D(N_i) + e(\overline{N_i, M_i};F)$, replace $D(M_i)$ by this smaller quantity and record $\overline{N_i, M_i}$ as the approach link associated with node M_i.

(3) If neither of the foregoing inequalities holds, make no change.

Repeat step b until the stage is reached when 3 applies for every link (M_i, N_i) of the network.

It is shown below that this stage is necessarily reached in a finite number of steps, that at this time $D(B) = \min_C e(C;F)$ for all A/B chains C, and that the final set of approach links can be used to determine a specific effectively shortest A/B chain C, if $\min_C e(C;F) < \infty$, i.e., if F is nonmaximal.

DERIVATION OF PROPERTIES OF THE ALGORITHM

The next five results (lemmas 9 to 13) state certain relations that exist between finitely labeled nodes and approach links at any stage of the algorithm. Assume that F is an ideal A/B flow pattern in a network and that after starting to apply the rules of the algorithm the process is interrupted at an arbitrary time. The labels and approach links described in the statements and proofs of these lemmas are assumed to be those that exist at the time of interruption except where otherwise noted.

Lemma 9: If $\overline{M, N}$ is an approach link, the following relation holds:

$$D(N) \geq D(M) + e(\overline{M, N};F)$$

Proof: Let $\bar{D}(M)$ denote the label of node M at the time it was used to assign the present label to node N. Then

$$D(N) = \bar{D}(M) + e(\overline{M, N}; F)$$

as a result of the relabeling rules. But $\bar{D}(M) \geq D(M)$ since labels are never increased by the process. The inequality of the lemma follows.

Lemma 10: The set of approach links do not contain a subset that is a loop.

Proof: Suppose $\overline{N_1, N_2, \ldots, N_k}$ $(N_k = N_1)$ is a loop, where each $\overline{N_i, N_{i+1}}$ is an approach link. This loop is necessarily a cycle. (Since a node never has two approach links associated with it, the terminal nodes of the links are all distinct.) According to lemma 9 then

$$D(N_i) \geq D(N_{i-1}) + e(\overline{N_{i-1}, N_i}; F)$$

for $i = 2, 3, \ldots, k$. Moreover strict inequality holds for at least one value of i. Let N_j be the first of these $k - 1$ nodes to attain its present label. Then if $\bar{D}(N_{j-1})$ denotes the label associated with N_{j-1} at the time N_j attained its present label, $\bar{D}(N_{j-1}) \neq D(N_{j-1})$ and thus $\bar{D}(N_{j-1}) > D(N_{j-1})$ since labels are nonincreasing. Thus

$$D(N_j) > D(N_{j-1}) + e(\overline{N_{j-1}, N_j}; F)$$

Adding the $k - 1$ inequalities, at least one of which is strict, and noting that $N_1 = N_k$,

$$\sum_{i=2}^{k} D(N_i) > \sum_{i=2}^{k} D(N_i) + \sum_{i=2}^{k} [e(\overline{N_{i-1}, N_i}; F)]$$

But then the cycle $\overline{N_1, N_2, \ldots, N_k}$ has negative effective length relative to F. This contradicts lemma 8. So no loop can exist, and the proof is complete.

Lemma 11: If $D(N) < \infty$ for a node N other than A, an A/N chain C exists whose links are all approach links.

Proof: Since N has finite label, an approach link $\overline{N_2, N}$ is associated with it. But then N_2 has a finite label and an approach link $\overline{N_3, N_2}$. One can continue to trace backward in this manner until at some stage an approach link $\overline{N_k, N_{k-1}}$ is reached such that $N_k = A$. For according to lemma 10 the same node is never reached twice, and there are a finite number of nodes. Thus $C = \overline{A, N_{k-1}, N_{k-2}, \ldots, N_2, N}$ is the desired chain.

Lemma 12: Node A retains its original label of zero.

Proof: Suppose node A is assigned a label $D(A) = n < 0$. Let $\overline{N, A}$ be the approach link associated with A. Then $D(N) < \infty$, and an A/N chain of approach links $\overline{A, N_{k-1}, N_{k-2}, \ldots, N_2, N}$ can be produced according to lemma 11. But then $\overline{N, A, N_{k-1}, \ldots, N_2, N}$ is a loop, which contradicts lemma 10. This completes the proof.

Lemma 13: If $D(N) < \infty$ the A/N chain C constructed in Lemma 11 has the property $D(N) \geq e(C;F)$. It follows that the label associated with a node is never less than the minimum effective distance from A to that node.

Proof: Let C be denoted by $N_1, N_2, \ldots, N_r$, where $N_1 = A$ and $N_r = N$. Then for $i = 2, 3, \ldots, r$ the following inequality holds according to lemma 9:

$$D(N_i) \geq D(N_{i-1}) + e(\overline{N_{i-1}, N_i}; F)$$

Adding these inequalities the following is obtained:

$$\sum_{i=2}^{r} D(N_i) \geq \sum_{i=1}^{r-1} D(N_i) + e(C;F)$$

Noting that $D(N_1) = 0$ according to lemma 12, the desired result follows.

Now it can be shown that all labels remain constant after a certain number of repetitions of step b (see the subsection "Statement of the Algorithm").

Lemma 14: All node labels remain constant after at most $n - 1$ iterations of (b) where n is the number of nodes in the network.

Proof: Let N be any node other than A such that $\min [e(C;F)] < \infty$, where the minimum is taken over all A/N chains C. Let $C_N = \overline{N_1, N_2, \ldots, N_k}$ be any A/N chain that attains this minimum. Clearly $k \leq n$ since no chain contains the same node twice. So C_N has at most $n - 1$ links.

After the first application of step b to all links, $D(N_2) \leq D(N_1) + e(\overline{N_1, N_2}; F)$ is obtained since N_2 is relabeled on the basis of $D(N_1) + e(\overline{N_1, N_2}; F)$ unless an even smaller label can be assigned. In general on completion of the i^{th} application of step b

$$D(N_{i+1}) \leq D(N_i) + e(\overline{N_i, N_{i+1}};F)$$

for $i = 1, 2, \ldots, k-1$, where $D(N_{i+1})$ is the label of N_{i+1} at the end of the i^{th} (or beginning of the $i+1^{st}$) application of step b and $D(N_i)$ is the label of N_i at the beginning of the i^{th} application. Adding these $k-1$ inequalities

$$\sum_{i=2}^{k} D(N_i) \leq \sum_{i=1}^{k-1} D(N_i) + e(C_N;F)$$

Since $D(N_1) = 0$ it follows that

$$D(N_k) \leq e(C_n;F)$$

On the other hand $D(N_k)$ cannot be less than the minimum effective distance from A to N, since according to lemmas 11 and 13 an A/N chain C of approach links can be produced so that $D(N_k) \geq e(C;F)$. It follows that C is an A/N chain of minimum length, and that

$$D(N_k) = e(C;F)$$

Thus after at most $k-1$ applications of step b, D(N) attains its minimum value. [For nodes N such that $\min e(C;F) = \infty$, D(N) is minimized from the start of the algorithm, so that the statement applies to all nodes of the network.] This completes the proof.

The following theorem summarizes the foregoing results.

Theorem 4: If the effective-length algorithm is applied to a network with an ideal A/B flow pattern F, the process terminates (i.e., all labels remain constant) after at most n - 1 repetitions of step b. Upon termination the final value of D(N) for any node $N \neq A$ is equal to min [e(C;F)], where the minimum is taken over all A/N chains C. Moreover if $D(N) < \infty$ a chain attaining this minimum can be found by tracing back along approach links, as in the proof of lemma 11.

Thus the effective-length algorithm possesses the properties that were assumed in the previous section. As a by-product it determines the effective distance from A to every other node, although only the effective distance from A to B can be used in the procedure.

Note that the sequencing of the algorithm, whereby the links are scanned in some fixed order in step b, is not claimed to be as efficient as possible. Pollack and Wiebenson[6] summarize a number of efficient labeling algorithms whose objective is to minimize l(C) over all A/B chains C, and it appears that any of these might be adapted to minimize e(C;F). These are generally sequenced in such a way that all links terminating in a given node are scanned together, and the nodes are taken in a certain order. The authors of this paper are not in a position to evaluate the relative efficiencies of the various schemes if programed for a digital computer.

GLOSSARY OF PRINCIPAL SYMBOLS

Symbol	Meaning
(M, N)	Link joining nodes M and N (considered undirected)
$\overline{M, N}$	Directed link from M to N
$\overline{M_1, M_2, \ldots, M_k}$	An M_1/M_k path if $M_1 \neq M_k$; a loop if $M_1 = M_k$
$c(\overline{M, N})$	Capacity of $\overline{M, N}$ (a nonnegative integer)
$l(\overline{M, N})$	Length of $\overline{M, N}$ (a nonnegative real number)
$l(\overline{M_1, M_2, \ldots, M_k})$	Length of path or loop
$f(\overline{M, N})$, $f(\overline{N, M})$	Flow in (M, N) expressed relative to $\overline{M, N}$ and $\overline{N, M}$ respectively
$Y(M)$	Net output at node M (flow output minus input)
$V(F)$	Value of A/B flow pattern F[equal to $Y(A)$ and $-Y(B)$]
$F \oplus G$	Sum of A/B flow patterns F and G
$F \ominus G$	Difference between A/B flow patterns F and G
$r(\overline{M_1, M_2, \ldots, M_k})$	The flow pattern obtained by assigning r units of flow to each $\overline{M_i, M_{i+1}}$ and zero to other links

$T(F)$	"Cost" of flow pattern F (length times flow, summed over all links)
$e(\overline{M, N}; F)$	Effective length of $\overline{M, N}$ relative to flow pattern F
$e(C; F)$	Effective length of chain or cycle C relative to flow pattern F
$D(M)$	Label assigned to node M by effective length algorithm.

REFERENCES

1. G. B. Dantzig and D. R. Fulkerson, "Computation of Maximal Flows in Networks," The RAND Corporation, P-677, 1 Apr 55.

2. D. R. Fulkerson, "On the Equivalence of the Capacity-Constrained Transshipment Problem and the Hitchcock Problem," The RAND Corporation, RM-2480, 13 Jan 60.

3. L. R. Ford, Jr., and D. R. Fulkerson, "Constructing Maximal Dynamic Flows From Static Flows," Ops. Res., 6: 419-33 (1958).

4. L. R. Ford, Jr., "Network Flow Theory," The RAND Corporation, P-923, 14 Aug 56.

5. L. R. Ford, Jr., and D. R. Fulkerson, "A Simple Algorithm for Finding Maximal Network Flows and an Application to the Hitchcock Problem," The RAND Corporation, RM-1604, 29 Dec 55.

6. M. Pollack and W. Wiebenson, "Solutions of the Shortest-Route Problem — A Review," Ops. Res., 8: 224-30 (1960).